Three–Way Multidimensional Scaling: Formal Properties and Relationships Between Scaling Methods

Sabine Krolak–Schwerdt

Department of Psychology, Saarland University, D-66041 Saarbrücken, Germany

Abstract. This paper is concerned with methods of three-way two-mode multidimensional scaling which were developed for the joint analysis of a number of promities matrices. The classification of these methods into trilinear and quadrilinear models (Kruskal (1983)) is outlined, and it is shown, that a number of specific properties and interpretations are associated with this classification the methods within each class have in common. Finally, relationships of the methods within and between the two model classes are outlined.

1 Introduction

Models and methods of three–way two–mode multidimensional scaling were developed for the analysis of individual differences in the representation of proximities. In the following, the scalar product form of the models will be outlined. The input data consist of a three–way data matrix $X = (x_{ijj'})$, $i = 1, \ldots, I$, $j, j' = 1, \ldots, J$, that can be thought of as comprising a set of $I(\geq 2) J \times J$ scalar products matrices. X_i, a slice of the three–way matrix, consists of estimated scalar products between objects j, j' and they represent the co–occurrence of objects under the point–of–view of an individual or a condition i. Obtained proximities estimated by a specific individual i may be transformed into scalar products by the procedures described by Torgerson (1958).

Three–way scaling models may be classified into those which rest on a trilinear or a quadrilinear decomposition of the data (Kruskal (1983)). Associated with the type of decomposition are different assumptions on the nature of individual differences.

2 Trilinear scaling methods

Trilinear approaches assume that there is a common space which underlies the objects in general. Thus, different subjects are presumed to perceive or judge stimuli on common sets of dimensions. On the basis of the common object space, it is assumed that individuals may distort the object space in their perception by attaching different importance or weights to the object dimensions than others do. The basic equation is given by

$$x_{ijj'} = \sum_{f=1}^{F} a_{if} b_{jf} b_{j'f} + e_{ijj'} \tag{1}$$

where F denotes the number of dimensions, or, equivalently,

$$X = A_F I(B'_F \otimes B'_F) + E \tag{2}$$

where I denotes the three–way identity matrix, A_F is a $I \times F$ matrix with the weights or saliences of the subjects on the F dimensions, reflecting the subject space, and B_F is a $J \times F$ matrix representing the common object space. The $e_{ijj'}$ are the errors of approximation collected in the three–way matrix E and $\otimes$ denotes the Kronecker product.

A very compact representation of trilinear models is obtained if the values of a subject i in the subject dimensions are arranged in a diagonal $F \times F$ matrix A_i. This yields the expression

$$X_i = B_F A_i B'_F + E_i$$

where A_i contains the dimensional weights under the point–of–view of subject i. Applying the individual weights to the common object space B_F yields the object space specific to the individual, $Y_i = B_F A_i^{\frac{1}{2}}$.

The most prominent trilinear model is the INDSCAL approach proposed by Carroll and Chang (1970). To fit the model to a given three–way data matrix, an alternating–least–squares algorithm is used.

Recently, two methods termed "SUMM–ID 1" and "SUMM–ID 2" (Krolak-Schwerdt (in press); Krolak-Schwerdt (1991)) were introduced where SUMM–ID 1 rests on the same model equation as INDSCAL, but uses another method to decompose the data matrix according to Equation (1). Basically, the method uses a three–way generalization of the centroid approach such that dimensions correspond to the centroids of the data. The central feature of the method is to base the object dimension b_f on the introduction of sign vectors z_f for the objects j, $z_{jf} \in \{-1, 1\}$, and, in an analogous way, to base the subject dimension a_f on sign vectors s_f for individuals i, $s_{if} \in \{-1, 1\}$, where

$$\sum_i \sum_j \sum_{j'} s_{if} z_{jf} z_{j'f} x_{ijj'} = \gamma_f := max \; . \tag{3}$$

The vectors z_f and s_f are the basis for the determination of dimension a_f and b_f in the following way:

$$\begin{aligned} a_{if} &= \frac{1}{\sqrt[3]{\gamma_f}^2} \; u_{if} \, , && \text{where} && u_{if} = \textstyle\sum_j \sum_{j'} z_{jf} z_{j'f} x_{ijj'} \, , \\ b_{jf} &= \frac{1}{\sqrt[3]{\gamma_f}^2} \; q_{jf} \, , && \text{where} && q_{jf} = \textstyle\sum_i \sum_{j'} s_{if} z_{j'f} x_{ijj'} \, , \\ & && \text{and} && \gamma_f = \textstyle\sum_i s_{if} u_{if} = \sum_j z_{jf} q_{jf} \; . \end{aligned} \tag{4}$$

In the equations, the scalar γ_f is a normalizing factor. After the determination of a_f and b_f in this way, the procedure continues in computing the residual data $x^*_{ijj'} = x_{ijj'} - a_{if} b_{jf} b_{j'f}$ and repeating the extraction of dimensions according to (3) and (4) on the residual data until a sufficient amount of the variation in the data is accounted for by the representation.

As Equation (3) shows, the values of the sign vectors z_f and s_f are determined in such a way that the sum of the data values collected in a dimension will be maximized. To compute the sign vectors, an algorithm is used that alternates between the subject and the object mode. That is, given the sign vector of one mode, multiplying the data matrix (the residual data matrix, respectively) with this sign vector yields the signs of the other vector. This is due to the cyclical relation between the modes

$$\begin{aligned}\textstyle\sum_i \sum_j \sum_{j'} s_{if} z_{jf} z_{j'f} x_{ijj'} &= \textstyle\sum_i s_{if} u_{if} = \sum_i \mid u_{if} \mid \\ &= \textstyle\sum_j z_{jf} q_{jf} = \sum_j \mid q_{jf} \mid = \gamma_f \ ,\end{aligned}$$

which also motivates the definition of the normalizing factor γ_f in Equation (4). The relation between the modes is used in the SUMM–ID algorithm by iteratively fixating one vector and estimating the other vector until the values of both sign vectors have stabilized. The procedure was first introduced by Orlik (1980).

3 Quadrilinear scaling models

Quadrilinear models share the assumption of the common object space with trilinear approaches. Furthermore, they assume that individual representations of the object space are distortions of the common object space. However, these distortions are more complex than in trilinear models. Besides the introduction of differential weights attached to the object dimensions, it is assumed that individual representations may be rotated versions of the common object space in which independent dimensions may become correlated.

The basic equation of the approaches can be expressed as

$$x_{ijj'} = \sum_{m=1}^{M} \sum_{p=1}^{P} \sum_{p'=1}^{P} a_{im} b_{jp} b_{j'p'} g_{mpp'} + e_{ijj'} . \tag{5}$$

A matrix formulation of the model is

$$X = AG(B' \otimes B') + E. \tag{6}$$

The coefficients a_{im} and b_{jp} are the elements of matrices A and B, the $g_{mpp'}$ are the elements of a three–way core matrix G (cf. Tucker (1972)). A is a $I \times M$ matrix with the coefficients of the subjects on M dimensions. B is a $J \times P$ matrix specifying an object space which is common to all subjects.

A more compact matrix representation is

$$X_i = BH_iB' + E_i \tag{7}$$

where the H_i, termed 'individual characteristic matrix' (cf. Tucker (1972)), is a linear combination of the M frontal slices, G_m, of the core matrix

$$H_i = \sum_{m=1}^{M} a_{im} G_m \tag{8}$$

Thus, H_i is a $P \times P$ symmetric matrix designating the nature of individual $i's$ distortion of the object dimensions. Diagonal elements h_{ppi} of H_i correspond to weights applied to the object dimensions by individual i, while off–diagonal elements $h_{pp'i}$ indicate the perceived relationships among the object dimensions p and p' under the point–of–view of individual i. As Equation (7) shows, the matrix H_i transforms the common object space into the individual representation.

The most prominent model of this type is the Tucker (1972) model which uses normalized principal components of scalar products of the data to derive the dimensions both in the subject space and in the object space. More precisely, writing the three–way data matrix as an ordinary two–way matrix $X_{(I)}$, $X_{(I)} \in R^{I \times JJ}$, by making use of combination modes (Tucker (1966)), A consists of the eigenvectors of $X_{(I)}X'_{(I)}$. In an analogous way, B consists of the eigenvectors of $X_{(J)}X'_{(J)}$, $X_{(J)} \in R^{J \times IJ}$. Kroonenberg and De Leeuw (1980) developed an alternating–least–squares algorithm to fit the Tucker model to a given three–way data matrix.

SUMM–ID 2 also rests on Equation (5). The model derives from its trilinear counterpart SUMM–ID 1 in the following way. It involves a rotation of the common object space

$$B = B_F T \ , \tag{9}$$

where the orthonormal transformation matrix T evolves from the singular-value decomposition $B_F = K \Delta_B^{\frac{1}{2}} T'$.

Analogously, the subject space is rotated

$$A = A_F V \tag{10}$$

with the orthonormal transformation matrix V deriving from the singular–value decomposition $A_F = L \Delta_A^{\frac{1}{2}} V'$. Inserting the rotated matrices into the model equation from SUMM–ID 1, that is

$$\begin{aligned} X &= A_F \quad I \quad (B'_F \otimes B'_F) + E \\ &= A \quad V'I(T \otimes T) \, (B' \otimes B') + E \\ &= A \quad G \quad (B' \otimes B') + E \end{aligned}$$

yields the final representation that was introduced in Equation (6).

As a very general approach, IDIOSCAL (Carroll and Chang (1970), Carroll and Wish (1974)) introduces a symmetric positive definite matrix C_i into the model equation $X_i = B \ C_i B'$ in order to allow for idiosyncratic rotations of the object space for different individuals. The central question within the model is how to decompose C_i. Two different ways of decomposing C_i have been proposed (cf. Carroll and Wish (1974)).

One procedure is given by $C_i = T_i \Delta_i T'_i$, with T_i orthonormal and Δ_i a diagonal matrix. Geometrically, the decomposition consists of an orthogonal rotation to a new coordinate system for subject i and a rescaling or weighting of dimensions of this new coordinate system by the diagonal entries of Δ_i. By this procedure, IDIOSCAL becomes equivalent to the SUMM–ID 2 formulation with the extra constraint in IDIOSCAL that individual spaces for subjects consist of orthogonal object dimensions. The second procedure is the one introduced by Tucker (1972). In this case, IDIOSCAL becomes an identical account to the Tucker model (cf. Kroonenberg (1994), Carroll and Wish (1974)).

4 Properties of trilinear and quadrilinear models

Associated with the classification into tri- and quadrilinear decompositions are specific properties the models within each class have in common. Perhaps the most important feature of trilinear models which is common to INDSCAL and SUMM–ID 1 is the 'intrinsic axis property' (Kruskal (1983)). That is, the dimensions of the object space B_F and of the subject space A_F are uniquely determined up to a joint permutation of the columns of the two matrices, and up to a scaling of the columns of the two matrices. Thus, both accounts provide solutions where the rotational position of dimensions is fixed.

Another special feature of trilinear models is that the number of dimensions in the subject space and the object space must be the same. In other words, one set of dimensions is extracted from the data matrix which then specifies the dimensionality F in both spaces.

Furthermore, both in INDSCAL and in SUMM–ID 1 the dimensionality of A_F and B_F is very high. For data forming a three–way array of size $I \times J \times J$ the number of dimensions necessary to reproduce the data is larger than $min(I, J)$ (cf. Kruskal (1976, 1983)). This implies that the dimensions are generally oblique and may become even linear dependent.

With respect to formal characteristics, quadrilinear models are different in nature. As has already been stated, these models introduce additional parameters by means of the core matrix G which have to be estimated along with the other unknown parameters in A and B. Thus, these models are more general formulations than trilinear approaches. As a consequence of using singular–value–decomposition either to extract dimensions or to rotate spaces, quadrilinear models retain rank properties of one–mode methods. That is, the number of dimensions does not exceed the number of subjects or objects and dimensions are always orthogonal.

Furthermore, as another consequence of the introduction of the core matrix, both the subject space and the object space are subject to rotations. Specifically, postmultiplication of the matrices A and B by orthonormal transformation matrices does not affect the model estimates provided that the core matrix is counter–rotated. Thus, these models are not uniquely identified. Another property due to the introduction of the core matrix is that the dimensionality M of the subject space may differ from the one of the object space which is P.

5 Connections between three–way scaling methods

As has been outlined above, trilinear and quadrilinear methods have very different properties. Thus, connections between these models have been found under rather restrictive conditions only. That is, the Tucker model becomes algebraically equivalent to INDSCAL if the core matrix can be diagonalized (Kroonenberg (1983), Carroll and Wish (1974)). However, there exist more general relations in that the quadrilinear models can be derived from trilinear methods by the SUMM–ID approach even in the case of unconstrained core matrices. These model connections will be outlined in the following.

SUMM–ID has some very close model relations to INDSCAL and the Tucker approach. The relationship to INDSCAL refers to SUMM–ID 1 and will be presented first in the following. Subsequently, the connections of SUMM–ID 2 to the Tucker model will be outlined.

To present the relationship between SUMM–ID 1 and INDSCAL, the triple product of the matrices A_F and B_F will be denoted in a more compact manner (cf. Kruskal (1976)) by $[A_F, B_F, B_F]$, that is $A_F I(B'_F \otimes B'_F) = [A_F, B_F, B_F]$. In the following, the SUMM–ID 1 will be termed $[A_F, B_F, B_F]$ and the INDSCAL representation will be referred to by $[\overline{A}, \overline{B}, \overline{B}]$. Theorem 1 states the relationship between the two models (cf. Krolak-Schwerdt (1991)).

Theorem 1. Suppose that $X = [\overline{A}, \overline{B}, \overline{B}]$ is of minimum rank R such that $R = rank(X)$. Then $R = F$, and there exist two permutation matrices P_A and P_B and two non–singular diagonal matrices D_A and D_B such that

$$\overline{A} = A_F P_A D_A \quad \text{and} \quad \overline{B} = B_F P_B D_B \quad \text{where}$$

$$[P_A D_A, P_B D_B, P_B D_B] = I.$$

As Theorem 1 states, the number of dimensions F of the SUMM–ID 1 representation is the same as the dimensionality of the INDSCAL representation. Furthermore, from the SUMM–ID 1 configuration the INDSCAL dimensions in both the object and the subject space may be derived by simply permuting and rescaling the dimensions where the effects of permutation and rescaling in the two modes compensate each other.

If the standard procedure of INDSCAL to scale the dimensions (that is, to normalize the object space so that the variances of projections of objects on the several dimensions are equal to one and to compensate the normalization in the object space by multiplying the weights in the subject space by the reciprocal scaling factors) is applied to the SUMM–ID 1 representation in this way, then SUMM–ID 1 and INDSCAL simply differ in the sequence in which the dimensions are extracted. Thus, both accounts create equivalent representations although they use a very different rationale and method of analysis. Stated in other words, INDSCAL may be derived from SUMM–ID 1 by two classes of transformations, permutation and rescaling of dimensions.

If the classes of permissible transformations on SUMM–ID are extended to orthogonal rotations of the subject and the object space, then the Tucker (1972) model derives from SUMM–ID 2 as will be shown in the following.

The Tucker model will be denoted by $X = \overline{A}\,\overline{G}(\overline{B'} \otimes \overline{B'})$. Writing the three–way data matrix as an ordinary two–way matrix $X_{(I)}$, $X_{(I)} \in R^{I \times JJ}$, $\overline{A}$ thus consists of the eigenvectors of $X_{(I)}X'_{(I)}$ and $\overline{B}$ consists of the eigenvectors of $X_{(J)}X'_{(J)}$, $X_{(J)} \in R^{J \times IJ}$ (cf. Section 3).

For the SUMM–ID 2 model $X = AG(B' \otimes B')$, the three–way core matrix G will be represented as an ordinary two–way matrix in the two different forms $G_{(M)}$, $G_{(M)} \in R^{M \times PP}$, and $G_{(P)}$, $G_{(P)} \in R^{P \times MP}$.

To present the relationship between SUMM–ID 2 and the Tucker model, the SUMM–ID 2 representation is rescaled according to the reasoning of the Tucker model. That is, the object space matrix $B = K\Delta_B^{\frac{1}{2}}$ as well as the subject space matrix $A = L\Delta_A^{\frac{1}{2}}$ are normalized so that the variances of projections of objects (of subjects, respectively) on the dimensions are equal to one. This yields K in the object space and L in the subject space. To compensate the normalization, the core matrix is rescaled in a reciprocal manner, that is, $\Delta_A^{\frac{1}{2}}G_{(M)}(\Delta_B^{\frac{1}{2}} \otimes \Delta_B^{\frac{1}{2}}) := Z_{(M)}$ or $\Delta_B^{\frac{1}{2}}G_{(P)}(\Delta_A^{\frac{1}{2}} \otimes \Delta_B^{\frac{1}{2}}) := Z_{(P)}$. With this normalization, Theorem 2 states the relationship between SUMM–ID 2 and the Tucker model (cf. Krolak-Schwerdt (1991)).

Theorem 2. Suppose W_A consists of the eigenvectors of $Z_{(M)}Z'_{(M)}$ and W_B consists of the eigenvectors of $Z_{(P)}Z'_{(P)}$. Then

$$\overline{A} = L\,W_A\,, \quad \overline{B} = K\,W_B \quad \text{and}$$

$$\overline{G}_{(M)} = W'_A\,Z_{(M)}\,(W_B \otimes W_B).$$

As Theorem 2 states, the Tucker model derives from SUMM–ID 2 by two transformations which consist of rescaling the SUMM–ID configuration and subsequently rotating the configuration orthogonally. The rescaling step involves normalizing the length of dimensions in the subject and in the object space as well as weighting the elements of the core matrix with the corresponding variances of the dimensions. The eigenvectors of the weighted core matrix constitute orthogonal rotation matrices which map the SUMM–ID 2 dimensions onto the Tucker configuration in both modes. Furthermore, the Tucker core matrix is obtained from orthogonally counter–rotating the rescaled SUMM–ID 2 core matrix.

6 Concluding remarks

Although differing in formal properties, there exist some general relationships between trilinear and quadrilinear models for three–way data analysis. The latter methods can be derived from the trilinear class without restricting the core matrix to a diagonal form.

In sum, SUMM–ID appears as a unifying account which establishes these connections and which may easily emulate other three–way multidimensional scaling representations, each by two transformations. To derive INDSCAL, the SUMM–ID 1 representation must be rescaled and dimensions must be permutated. To obtain the Tucker (1972) model, the rescaling of the SUMM–ID 2 representation must be followed by orthogonal rotations of the configuration. From this it is directly evident, that IDIOSCAL may be obtained from SUMM–ID 2.

Finally, it should be noted on more practical grounds that the estimates from SUMM–ID are computationally quite efficient and fast to obtain in terms of CPU time as compared to alternating–least–squares algorithms.

Another aspect of importance in applied research concerns the interpretation of the core matrices. Within the Tucker model the eigenvalues of dimensions obtained in both modes are connected with the core matrix which makes the interpretations of the values of the core matrix rather difficult (cf. Kroonenberg (1983)). In contrast, the core values in SUMM–ID 2 were defined as

$$g_{mpp'} = v_m' I(t_p \otimes t_{p'}) = [v_m t_p t_{p'}]$$

with unit length vectors v_m, t_p and $t_{p'}$. This corresponds to the three–way generalization of the standard scalar product. Thus, in SUMM–ID 2 values of the core matrix have a much more specific meaning in terms of interrelations between dimensions which makes the model parameters more easy to interpret.

References

CARROLL, J.D. and CHANG, J.J. (1970): Analysis of individual differences in multidimensional scaling via an N–way generalization of Eckart–Young decomposition. *Psychometrika, 35, 283–319.*

KROLAK-SCHWERDT, S. (1991): Modelle der dreimodalen Faktorenanalyse: formale Eigenschaften, theoretische Zusammenhänge und ihre Implikationen für das Konzept individueller Differenzen. *Psychologische Beiträge, 133, 314–346.*

KROLAK-SCHWERDT, S. (in press): A three-way multidimensional scaling approach to the analysis of judgments about persons. *Studies in Classification, Data Analysis, and Knowledge Organisation, Vol. 27 Classification: The ubiqitous challenge.* Springer, Berlin.

KROONENBERG, P.M. (1983): *Three–mode principal component analysis.* DSWO Press, Leiden.

KROONENBERG, P.M. (1994): The TUCKALS line. A suite of programs for three–way data analysis. *Computational statistics and data analysis, 18, 73–96.*

KROONENBERG, P.M. and DE LEEUW, J. (1980): Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika, 45, 69–97.*

KRUSKAL, J.B. (1976): More factors than subjects, tests and treatments: An indeterminacy theorem for canonical decomposition and individual differences scaling. *Psychometrika, 41, 281–293.*

KRUSKAL, J.B. (1983): Multilinear methods. In R. Gnanadesikan (Ed.): *Proceedings of Symposia in Applied Mathematics* (Vol. 28, pp. 75-104). Providence: American Mathematical Society.

ORLIK, P. (1980): Das SUMMAX-Modell der dreimodalen Faktorenanalyse mit interpretierbarer Kernmatrix. *Archiv für Psychologie, 133, 189–218.*

TORGERSON, W.S. (1958): *Theory and methods of scaling.* Wiley, New York.

TUCKER, L.R. (1966): Some mathematical notes on three-mode factor analysis. *Psychometrika, 31, 279–311.*

TUCKER, L.R. (1972): Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika, 37, 3–27.*